Superiority of a Handheld Perspective-Coupled Display in Isomorphic Docking Performances

Thibault Louis, Francois Berard
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG
F-38000 Grenoble France
(thibault.louis,francois.berard)@univ-grenoble-alpes.fr

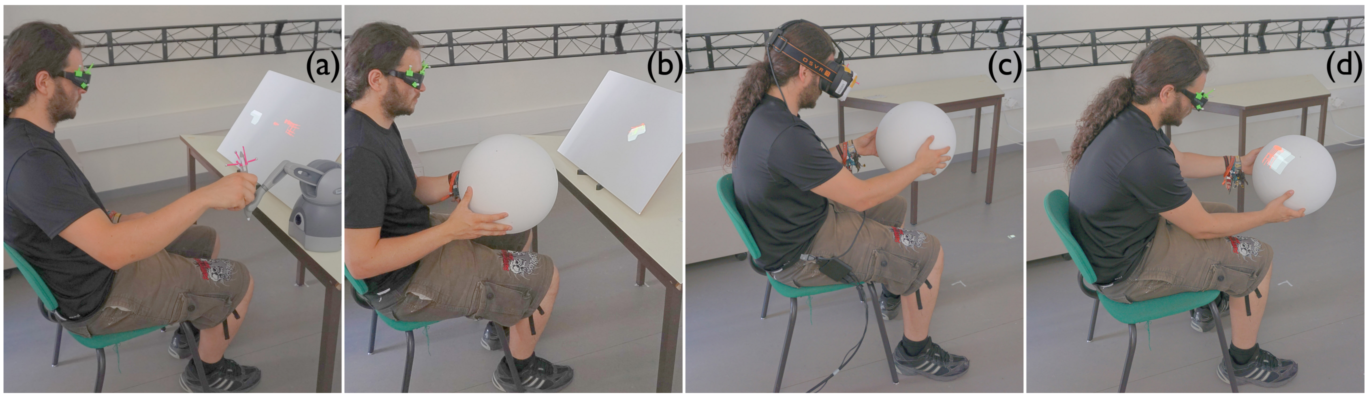

Figure 1. The experimental conditions. a) articulated arm (PHANTOM). b) tangible prop (PROP). c) head-mounted display (HMD). d) handheld perspective-coupled display (HPCD).

ABSTRACT
Six degrees of freedom docking is one of the most fundamental tasks when interacting with 3D virtual worlds. We investigated docking performances with isomorphic interactions that directly relate the 6-dof pose of the input device to that of the object controlled. In particular, we studied a Handheld Perspective-Coupled Display (HPCD); which is a novel form of interactive system where the display itself is handheld and used as the input device. It was compared to an opaque HMD and to a standard indirect flat display used with either a sphere or an articulated arm as the input device. A novel computation of an Index of Difficulty was introduced to measure the efficiency of each interaction. We observed superior performances with the HPCD compared with the other interactions by a large margin (17% better than the closest interaction).

Author Keywords
handheld perspective corrected display; 3D display; 6-dof docking; isomorphic control; evaluation; head-mounted display

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ISS '17, October 17–20, 2017, Brighton, United Kingdom
© 2017 Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-4691-7/17/10$15.00
DOI: **https://doi.org/10.1145/3132272.3134124**

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces - Input devices and strategies; I.3.7 Computer Graphics: Three-Dimensional Graphics and Realism - Virtual reality

INTRODUCTION
Docking an object in 3D space corresponds to moving the object to a particular position and orientation. It is a 6-degrees of freedom (dof) task (a combination of a 3-dof positioning and a 3-dof rotation of the object). Docking is often seen as one of the most important and frequent task in the edition of 3D virtual scenes. Most professional 3D editing software implement docking with the help of the 2D mouse and a sequence of interactions that decompose a single docking task into several subtasks.

However, docking can also be executed in a single *coordinated* gesture. This kind of gestures is coming back into focus as the interest for Head-Mounted Displays (HMDs) was renewed recently due to significant technical improvements and costs reduction. HMDs, both the opaque ones (e.g. Oculus Rift, HTC Vive) and the see-through ones (e.g. Microsoft Hololens, Meta 2), render realistic 3D digital objects in the surroundings of the user. In this context, users typically interact with them using an *isomorphic* mapping: the 6-dof motion applied by the user to the input device is reproduced "as is" on the digital object under control. With this mapping, docking can be executed in a similar way as we move and orient objects in the physical world.

The choice of the input device and interaction technique used in an isomorphic control docking has a clear effect on users' performances, as shown in previous studies [18, 19, 23]. Here, we test a novel form of interactive system: Handheld Perspective Coupled Displays (HPCDs) [16]. We use the display itself as the input device for docking. A recent study demonstrated the feasibility of a lightweight spherical HPCD [2]. The authors reported that participants found the display very easy to move around and that they had excellent control on its orientation. While the study focused on the *perception* of a complex 3D object, it suggested that the HPCD could be used as en efficient input device for 6-dof control.

In this paper, we present an empirical study aimed at quantifying users' efficiency in coordinated docking using a spherical HPCD. We compare its performances with those achieved with 3 other interactions. Two interactions also use the sphere as the input device, but vary in the display: an opaque HMD and a common flat display used in an indirect interaction. The third interaction uses the flat display and an articulated arm (Phantom Omni) as the input device.

We contribute by showing a large improvement (17%) in users' docking performance using the spherical HPCD over to the second most efficient interaction (the HMD), and a 43% improvement over the articulated arm that was previously considered as one of the most efficient docking interaction. As a secondary contribution, we introduce a novel approach to quantify the difficulty of a 6-dof docking task. It unifies the tolerances in translation and rotation and allows the computation of a single *throughput* metric that we estimate for the 4 tested interactions.

In the following, after reviewing the literature on docking interactions and docking performance studies, we introduce our novel approach to estimating the difficulty of a docking. We then introduce our comparative study on docking performances with four different interactions. After presenting the study's result, we discuss the main factors that can explain the large differences in performance and we conclude by addressing possible follow-up studies.

RELATED WORK

Interaction Techniques for Docking

In their book on 3D user interfaces, Bowman et al. [5] explain that being able to move an object anywhere on a scene and to orient it represents one of the most fundamental tasks in the interaction with virtual environments. Allowing users to perform this 6-dof task has been the focus of many research efforts. Those were initially oriented towards creating new input devices with an increased number of degrees of freedom compared to the computer mouse [6, 20, 22]. These devices support the solving of the 6 dof of the task *simultaneously* in a *coordinated* manner. However, Masliah et al. studied coordination and found that solving all dof at the same time may have a motor control demand that is too high for users, and that users tend to decompose the task into subtasks of lower dof [11]. In addition to their high motor control demand, high dof devices don't afford the stability and precision offered by a desktop mouse. This could explain why the mouse was found more efficient than higher dof devices in a 3 dof task [1], and why it remains the input device of choice for 3D modeling, i.e. including docking. More recently, multitouch interaction has spurred the creation of novel interaction techniques that exploit several fingers to support 6-dof controls [7, 8, 10]. However, as with the mouse, these interactions require complex mappings to transform one or more 2 dof translations on a plane into the 6 dof of the controlled object.

In this paper, we focus on docking interactions that offer an isomorphic control of the object. An isomorphic control is not always the most efficient form of control. Poupurev et al. observed for example that a non-isomorphic control of 3 dof rotations was 13% more efficient than an isomorphic one [13]. However, they demonstrated that non-isomorphic mappings can break important usability properties of the interaction: directional compliance and nulling compliance. One of the benefits of isomorphic control is that it reproduces the gestures performed when interacting with physical objects. As such, it can be expected that our extensive manipulation skills, acquired in the physical world, transfer to the virtual environment, making it more "natural" in the sense of Bérard and Rochet-Capellan [3]. An isomorphic control is also a requirement to activate the use of one's proprioception. Mine et al. define proprioception as "a person's sense of the position and orientation of his body and limbs" and they make a strong case for the use of proprioception in virtual environments [12]: proprioception allows people to grab their keys in their pocket *eyes free*, a strong evidence of its utility for aimed movements.

Users' 6-dof Docking Performances

Ware presented an early experiment on docking using isomorphic mapping [20] but the difficulty of the task was not explored. He used a mid-air device with a button to activate the control. The control was thus *relative* and did not allow the benefit of proprioception. Later, Zhai et al. established the use of a tetrahedron in the evaluation of docking performances [23]. They tested two mid-air isomorphic interaction techniques and demonstrated the benefit of using the fingers when controlling the rotations of the input device. Froehlich et al. presented two new non-isomorphic input devices for docking that they tested alongside the spacemouse, a purely rate-control device [6]. They demonstrated the benefit of an isotonic position-based control of rotation. Hancock et al. introduced various non-isomorphic multi-touch interaction techniques and tested them on a docking task [7]. In a pilot study, Vuibert et al. found that using a more common 3D model such as a chair, instead of a tetrahedron, facilitated the perceptual complexity of a docking task [18]. They tested 4 isomorphic interactions, including an articulated arm that was found superior in terms of accuracy but not in term of time. Besancon et al. recently compared the 3 most common interaction paradigms for docking: desktop (mouse+keyboard), multi-touch, and mid-air isomorphic [4]. They found the isomorphic interaction as the fastest while all condition provided the same level of accuracy.

An important limit of these studies is that the task accomplishment time was never related to various levels of difficulty of the dockings. Zhai et al. did vary the amplitude of the task (not the tolerance) but they did not report how this influenced the accomplishment time [23]. As a consequence, it is difficult to compare the various interactions tested in these studies. We provide a first effort to define an Index of Difficulty (ID) for 6-dof docking, and we use it to estimate the information throughput of the tested interactions.

In addition, while the use of an HPCD for docking was envisioned by Spindler et al. [15], its efficiency was never formally tested. This work is a first contribution in this domain.

AN INDEX OF DIFFICULTY FOR 6 DOF DOCKING

Stoelen and Akin offered a first step towards combining the translation and rotation components in the index of difficulty (ID) of a docking task [17]. However, they only addressed 2×1D docking. Furthermore, the combination only concerned the computation of the ID, not the task itself: they use 2 separated targets for the translational and the rotational components of the task. We build on this first step to propose a new formulation of the index of difficulty that models a full 6-dof docking and combines the error in translation and in rotation for the task.

Error Tolerance

We use a single *spatial tolerance* for the translation and rotation error of a successful docking. We claim that this has more ecological validity than using 2 independent tolerances. To illustrate this claim, we consider a shapes box game for children in the physical world where a cube-shaped block must be inserted in a square hole. The size difference between the hole and the cube represent a single *spatial tolerance* for this task. The task is successful, i.e. the cube can be inserted, only if the cube's face is entirely within the boundary of the square hole *whatever the rotation and/or position error*. In particular, if the cube's translation makes it perfectly aligned with the hole's center, then the entire tolerance can be used for a rotation error, and conversely. In the more general case, the rotation and translation error add up, and the success of the insertion depends on this sum being smaller than the single spatial tolerance that we call ET.

To generalize ET to any shape of the object, we define the *target* of a docking task as the volume resulting from the expansion by ET of the volume used by the virtual object placed at the target position and orientation. A docking is successful when the object is entirely contained in the target volume, whatever its position and orientation error. This is illustrated in Figure 2.

Index of Difficulty

To express the Index of Difficulty (ID) of the task, we start from Stoelen and Akin's model. It computes the docking ID as the sum of an ID for translation and an ID for rotation (equ. 1, form [17]), where the IDs are expressed as in the Shannon formulation of the Fitts' law (equ. 2, from [14]).

$$ID_{combined} = ID_{translation} + ID_{rotation} \tag{1}$$

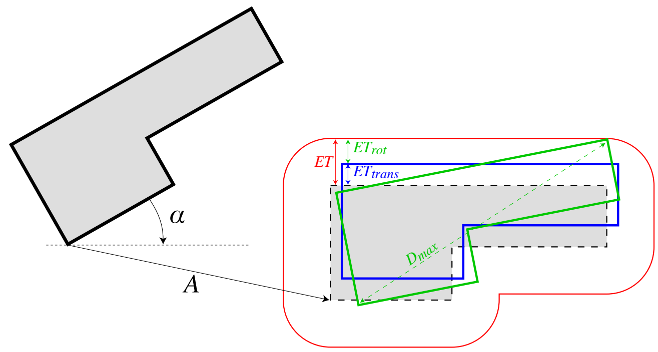

Figure 2. Parameters of a docking task of translation amplitude A and rotation amplitude α. The object (grey shape with a thick solid outline) must be docked to a target pose (grey shape with a dashed outline). The error tolerances in translation ET_{trans} and rotation ET_{rot} add up to the overall error tolerance ET. D_{max} is the largest distance between 2 points of the object, it allows the conversion from a rotation error to a spatial error (details in the paper).

$$MT = a + b.ID, \quad ID = \log_2\left(\frac{A}{W} + 1\right) \tag{2}$$

Fitts' law expresses the mean time MT to reach a target of index of difficulty ID, where ID only depends on the amplitude of the task A (i.e. the initial distance to the target) and the width of the target W. W is twice the error tolerance: success implies that the object is not further away than half the width from the target's center.

Equation 1 can be expanded using the amplitude and target width of the translation component (A, W) and of the rotation component (α, ω).

$$ID_{combined} = \log_2\left(\frac{A}{W} + 1\right) + \log_2\left(\frac{\alpha}{\omega} + 1\right) \tag{3}$$

In this equation, the amplitudes A and $alpha$ are simply defined as the translation and rotation distances between the object initial pose and target pose. The target widths W and $omega$ are more difficult to determine. They are linked by the following equation:

$$ET_{trans} + ET_{rot} = ET \tag{4}$$

where ET_{trans} and ET_{rot} are the spatial tolerances in translation and rotation, respectively. When expressed in terms of target widths, equation 4 translates to:

$$W + f(\omega) = 2 \times ET \tag{5}$$

where $f(\omega)$ translates the rotation's angular target width into a spatial tolerance. To evaluate this function, we make the assumption that users will solve the rotation component of the docking by rotating around the optimal axis: the one that minimalizes the amount of rotation. We use an *optimal rotation plan*, that is orthogonal to this axis, to project the points of the controlled object. The maximal distance D_{max} between two projected points is used to compute the largest *spatial error* resulting from an *angular error* ω. In other words, we compute $f(\omega)$ as follow:

$$f(\omega) = sin(\omega) \times D_{max} \quad (6)$$

Note that the equation 6 is only valid for angular target widths that are smaller than 90°, but this is not a constraint for docking tasks. In summary, for one particular docking task, there exists an infinity of possible IDs where W and ω can vary so that:

$$W + sin(\omega) \times D_{max} = 2 \times ET \quad (7)$$

Here, we make the assumption that the participant will unconsciously aim at the best combination of the translation and rotation target widths: the one that minimizes the difficulty of the task. Hence we numerically solve for the minimal value of $ID_{combined}$ according to equations 3 and 7.

The validity of this index of difficulty will be discusses in regard to the result of the empirical study.

EMPIRICAL STUDY

We ran an empirical study to measure participants' performances in 6-dof docking when using the 4 different isomorphic interaction techniques illustrated in Figure 1: an articulated arm, a tangible prop with a regular flat display, the same prop used with an opaque HMD, and a spherical HPCD. We name these 4 conditions PHANTOM, PROP, HMD and HPCD, respectively.

Technical Setup

In order to implement a spherical HPCD, we reproduced the setup described in [2] that we adapted to our specific needs. Various input and output devices have differences in performance such as precision and resolution; which can have an effect on users' performances. We tried to reduce these differences as much as possible by using the same sensing and display devices whenever possible. In particular, we did not use the native tracking of the HMD nor the one of the articulated arm. Tracking was implemented in all conditions using a 10-camera Optitrack optical tracking system running at 240 Hz and providing stable and accurate passive marker positions. This system was used to track both the participants' head (i.e. the HMD or stereo shutter glasses) and the input device. Constellations of markers were attached to the HMD, the stereo shutter glasses, the arm's pen, and the sphere, in order to recover their 6-dof pose in space.

Visual feedback was created for all conditions except HMD by a projector providing 2560×1600 pixels at 120 Hz. LCD shutter glasses were synchronized with the projector to provide stereo images at 60 Hz per eye. In the PHANTOM and PROP conditions, the image was formed on a planar screen made from a 600×375 mm white board (28", 16:10 aspect ratio, 82 dpi). In the HPCD condition, the image was formed on a lightweight white sphere made from a hollow polystyrene sphere having a diameter of 30 cm. Resolution was around 82 dpi, depending on the distance to the projector. The HMD condition used a Razer OSVR HDK 2 opaque HMD providing 1080×1200 pixels per eye at 90 Hz and a 110° field of view. Hence, the HMD condition had different graphical characteristics compared to the other conditions. The projector had an update rate of 120 Hz that was interleaved at 2×60 Hz on the two eyes. In terms of resolution, the pixel density in the HMD (approximately 32 dpi) was clearly lower than that of the projections: pixels are spread over the large field of view of the HMD; which results in pixel boundaries being perceptible in the foveal field of view.

The input device was the white sphere for PROP, HMD and HPCD. The sphere was held with both hands due to its large diameter. Being wireless and lightweight, it had no constraint in terms of workspace, although in the HPCD condition the range was constrained by the projection area. In the PHANTOM condition, we used a Geomagic Touch force feedback articulated arm. The arm was unpowered and did not generate any force feedback, except from its passive resistance that contributed to the stabilization of the attached pen. The device offers a workspace of 160×120×70 mm.

All conditions offered head-coupled stereoscopic graphical feedback.

Conditions

In the following, we refer to the *size of the scene* as the distance between to two most distant targets in the experiment (i.e. the diameter of the bounding sphere of all targets).

PHANTOM. We tested this condition seeing that it used an isomorphic docking technique that was recently found as efficient [19] and accurate [18]. Participants held a pen that was attached to the articulated arm, as illustrated in Figure 1a. The 6-dof motion of the tip of the pen was reproduced on the controlled object. We placed the device directly to the right of the screen (or to its left for left-handed participants). In pilot trials, we observed that providing physical support for the wrist or the forearm improved stability but created obstacles when acquiring the sequence of targets. We thus opted to leave as much free space to the participant's hand as possible. The size of the scene in *motor space* was 97 mm, which was one sixth of the size in the HMD and HPCD conditions. We used this scaling due to the limited workspace of the input device, and so that all targets could be acquired without any clutching. The graphical feedback was generated so that the center of the scene (the global center of all targets) appeared at the center of the flat display. The scene in the *display space* was zoomed-in by a factor three so as to maximize the use of the display surface.

PROP. We tested this condition in order to isolate the effect of the input device and the display: HMD and HPCD used the same input device but with different displays. Participants held the sphere between their hands to translate and rotate it. The 6-dof motion of the sphere was reproduced on the con-

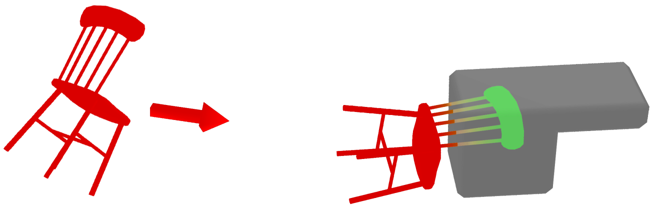

Figure 3. Graphical feedback for the task. This screenshot is edited for illustration purpose: the background is changed from black to white, the target from semitransparent white to gray, and 2 chairs are shown (a single chair was shown in the experiment). Parts of the chair that were inside the target turned green. A red arrow pointing at the target was only shown when the target was far away.

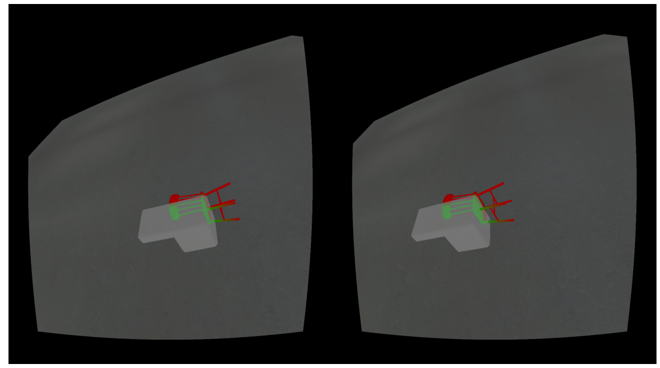

Figure 4. The scene as seen in the HMD. Only a textured ground plane was added in addition to the chair and target.

trolled object. Participants looked at the graphical feedback on the flat display, as illustrated in Figure 1b. The size of the scene in *motor space* was 290 mm, or half the size compared to the HMD and HPCD conditions. We used this scaling to avoid forcing participants into uncomfortable posture where the sphere would have to be held far to the side of the display while looking at it. We positioned the graphical scene in *display space* as with PHANTOM, but with no scaling factor as the size in motor space fitted well in the display surface.

HMD. Participants were equipped with the HMD and were given the sphere as illustrated in Figure 1c. In order to provide a contextual cue, we added a floor in the virtual scene that had a similar texture as the physical floor, as illustrated on Figure 4. We implemented a 1 to 1 mapping between the motor and display spaces: motions of the sphere were exactly replicated on the controlled object. The size of the scene was 580 mm. Participants were sat in front of the scene so that all targets were at arm's reach. This proximity to the scene required significant rotation around the torso axis for the head to see all targets and for the arms to reach them.

HPCD. Participants looked at the virtual scene "through" the sphere as illustrated in Figure 1d. We implemented to same 1 to 1 mapping between the motor and display space as with HMD, and we sat participants in a similar position at arm's reach from the targets. This implied the same head and arms rotation around the torso axis to cover all the targets.

On varying the scale of the scene. As described above, we scaled the scene in motor space in an effort to offer an optimal scale for each condition. HMD and HPCD benefit from a large scene to facilitate accurate position control while covering large distances is performed easily by fast translations of the sphere. This large scene is not suitable for PHANTOM, as it would require a lot of clutching to execute large translations. By varying the scale of the scene across conditions, it may appear that we favor the ecological validity of the experiment (conditions are tested at a realistic scale) over its internal validity; which suggests using the same scale across conditions. However, it is important to note that scaling of the scene changes both the amplitude *and* the target size in the same proportion, hence it does not change the task difficulty and task accomplishment time as modeled in equation 2. In addition, this scaling has no effect on the rotational component of the docking.

Task

The experimental task was modeled after the docking task introduced at the beginning of this paper. Our study was focused on the *motor control* qualities of the four tested interaction techniques and not on the difficulty of perceiving the mismatch between the object's current and target pose, hence we attempted to minimize the latter. The controlled object is a model of a chair as in Vuibert et al. [18]. The targets of the dockings are represented by white semitransparent surfaces resulting from the expansions of the chair by various spatial tolerances. The chair is displayed in red when outside the target. Parts of the chair that are inside the target turn green; the participants' goal is to turn the chair entirely green. This color-coding helps participants identify which parts of the chair are outside the target, and plan for a corrective motion. At any point in time, the virtual scene only contains the chair and the target volume. In addition, when the object is far away from the target, we show a red arrow near the object pointing to the target, as illustrated on Figure 3.

We wanted to observe how each technique could mobilize participants' proprioception to improve docking performances. Hence, we selected a limited number of targets (5) that were always presented in the same sequence. This way, participants quickly learned the sequence of targets and their poses in space. Once learned, participant would be able to solve a large part a docking with their eyes closed using only their proprioception (even though they kept their eyes opened in the experiment).

There was no validation required from the participants: a docking was validated as soon as the object was entirely within the target's boundary. This has the benefit of preventing unwanted motion caused by any form of validation, and the speed/accuracy tradeoff is implicitly normalized across participants (e.g. participants cannot spend time trying to be very accurate before validation). Furthermore, the contribu-

ID (bit/s)	6.0	8.0	10.0	12.0	13.5
ET (mm)	7.2	9.1	5.8	4.2	4.3
A (mm)	54	242	203	343	580
alpha (°)	23	40	80	108	180

Table 1. Indices of Difficulty (IDs) used in the experiment. They are computed by minimizing $ID_{combined}$ in the equations 3 and 7. A and ET are expressed for HMD and HPCD, they were scaled down for PHANTOM and PROP.

tion of the reaction time is reduced in the task accomplishment time because participant can anticipate the next target.

A secondary objective of the experiment was to test the validity of the Index of Difficulty (ID) for docking introduced at the beginning of the paper. We chose 5 targets that yielded a nice spread of IDs as detailed in table 1

Design

We welcomed 20 volunteer participants (4 women, 2 left handed, age average 27.4 [20..38]). We used the Stereo Optical RANDDOT stereopsis test to check that all participants had adequate stereovision. None of them had experience neither with an HMD nor the articulated arm. One participant was accustomed to virtual objects manipulation, and another one to stereoscopic 3D scene visualization. We checked that the results of these two participants were similar to the results of the others. After welcoming the participants, we provided an instruction sheet explaining the purpose of the experiment and the protocol.

We used a within-subject design. The order of presentation of the conditions was completely counterbalanced across participants. We instructed participants to be as fast as possible. When starting with a new condition, participants trained on a block of 5 repetitions of the five targets series (25 dockings), in order to familiarize with the new device. After training, they performed two successive measured blocks of 11 repetitions of the 5 targets. Between each blocks, participants could take a break if they wanted to. The first repetition in the two measured blocks was considered as a warm-up and was ignored. In summary, we recorded 20 participants $\times$ 4 conditions $\times$ 2 measured blocks $\times$ 10 repetitions $\times$ 5 targets = 8000 docking trials.

We tested 2 independent variables: TECHNIQUE and ID. We recorded the *docking_time* as the time between 2 successful dockings. In addition, all the tracking data was recorded. At the end of the session, participants filled a questionnaire to provide their subjective judgment on the interaction techniques. Overall, experimental sessions lasted around 45 minutes.

Results

We performed a repeated-measure ANOVA on *docking_time* with TECHNIQUE and ID as main factors. It revealed a strong effect of TECHNIQUE ($F(3, 57) = 47.2$, $p = 1.9e-15$), a strong effect of ID ($F(4, 76) = 88.6$, $p < 2e-16$), and a strong interaction between the two ($F(12, 228) = 3.85$, $p = 2.3e-05$).

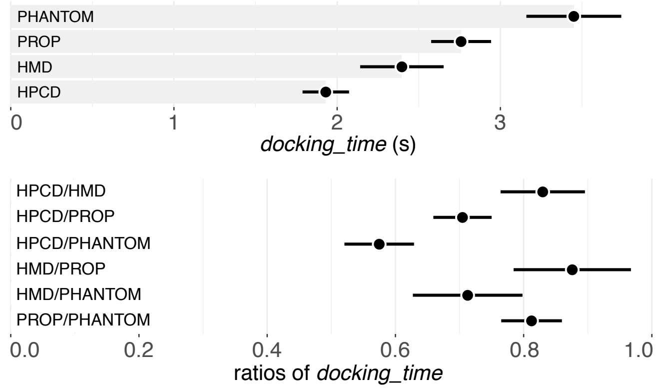

Figure 5. Top: *docking_time* in each condition, averaged across participants and with 95% confidence intervals. Bottom: participants' ratio of *docking_time* for all pairwise combination of conditions, averaged over participants and with 95% confidence intervals.

Condition	Slope	p-value	R-squared	Throughput bit/s
PHANTOM (including ID=6)	0.301	0.041	0.798	3.1
PHANTOM (without ID=6)	0.446	0.020	0.961	3.3
PROP	0.354	0.006	0.941	3.8
HMD	0.258	0.012	0.911	4.4
HPCD	0.207	0.004	0.953	5.4

Table 2. Parameters of linear regressions of *docking_time* predicted by ID for each condition.

Docking time

The average docking time per condition across participants and ID are 3.45 s, 2.76 s, 2.40 s and 1.93 s for PHANTOM, PROP, HMD and HPCD, respectively. Docking times are represented on Figure 5, top. We computed the participants' all pairwise variation of the docking time across conditions, this is represented on Figure 5, bottom. There was a clear separation of the performances of each technique: on average, participants improved their performances with HPCD by 17.0% over HMD, HMD improved 12.4% over PROP, and PROP improved 18.8% over PHANTOM. All pairwise t-tests with Holm correction revealed significant differences with $p <= 0.003$ in all cases except PROP vs. HMD with $p = 0.013$.

Fitts regressions

The strong interaction between TECHNIQUE and ID is a first indication that the ID can be used to compute an information throughput per technique, and thus to compare the efficiency of the techniques. For each technique, we plotted the *docking_time* according to ID and computed a linear regression. This is illustrated in Figure 6. The *docking_time* for the easiest ID in PHANTOM is clearly misaligned. This target required a 54 mm translation to the right and a 23° rotation around a horizontal axis parallel to the display. Performing this rotation with the tip of the pen proved difficult,

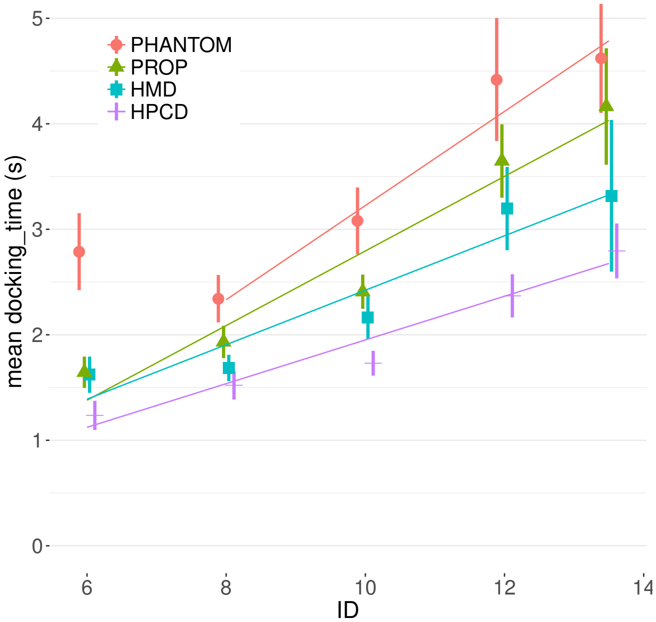

Figure 6. ***docking_time*** **per ID in each condition with 95% confidence intervals. Linear regressions are represented with ID=6 ignored for PHANTOM (details in the text).**

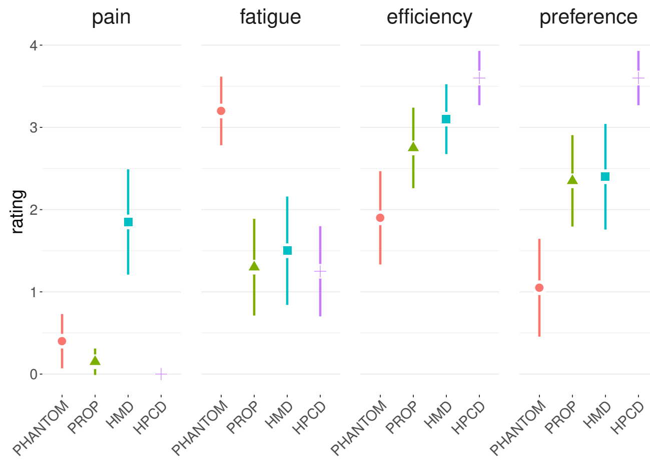

Figure 7. Subjective ratings with 95% confidence intervals.

while there was no particular difficulty with the unconstrained sphere in the other conditions. We report in table 2 the parameters of the linear regressions, with two sets from PHANTOM: with and without ID=6. Throughput was estimated per condition as the grand average of the ratio of ID over *docking_time* across all IDs and participants, as advocated in [14].

Subjective results

The post-experiment questionnaire measured participants' subjective evaluation of each technique according to 4 criteria. Participants were asked to rate 4 sentences per technique with a score ranging from 0 ("I totally disagree") to 4 ("I totally agree"). The sentences were: "This technique give me pain in the eyes, headaches, or nausea." (pain), "This technique is the source of muscular fatigue." (fatigue), "This technique allows me to position the chair efficiently (quickly and precisely)." (efficiency), "I like this technique." (preference). Results are summarized in Figure 7.

DISCUSSION

Validity of the Index of Difficulty

We introduced a novel way to compute an Index of Difficulty for a 6-dof docking task that models the sharing of a spatial tolerance between the translational and rotational error. A first sanity check for the validity of this ID is that docking time systematically increased with ID in each condition. A more ambitious objective was to use this ID to allow the measure of the information bandwidth in bit/s of each technique. This required that task accomplishment time vary linearly with ID. Although it can be seen in Figure 6 that the alignments are not perfect (with a tendency to overestimate at ID=10), the parameters of the linear fits reported in table 2 indicate that the fits are good and support a close to linear relationship. In turn, the goodness of the fits support the validity of the throughputs reported in the same table.

In their review of previous Fitts studies on 2D pointing, Soukoreff and Mackenzie report that IDs were tested in the range 1 to 6 (table 4 in [14]). We tested more difficult tasks in the range 6 to 13.5. This is to be expected as our 6-dof docking combines the difficulty of two 3-dof tasks which require much more information than a single 2-dof task. Nonetheless, throughput values represent the information bandwidth of the combination of a user and an interaction technique. They should be in the same order of magnitude whatever the task, assuming that the techniques are not critically inefficient. Soukoreff and Mackenzie found throughputs in the range 3.7 bit/s to 4.9 bit/s for studies of the standard 2D mouse. The throughputs estimated in this docking experiment, in the range 3.1 bit/s to 5.4 bit/s, are remarkably similar. At 5.4 bit/s, the throughput estimated for HPCD is similar to the best 2D mouse techniques reported in [14]. This indicates that HPCD makes a very good use of the information capacity of its users, although we should be careful in comparing throughput values estimated in different ways and on quite different tasks. Still, the results of this first effort on estimating the throughput in a 6-dof docking are encouraging, and should foster further studies and replications.

Difference in Efficiency Across Conditions

We took great care in the experiment to minimize the technical differences across conditions, and some of the conditions shared essential characteristics: PHANTOM and PROP used the same indirect display; PROP, HMD and HPCD used the same input device. Yet, performances were almost regularly spread across conditions, as illustrated in Figure 5, top. This indicates that the factors that we tested, display and input device, and their combinations, had an important influence on users' efficiency in docking.

There is a remarkable correspondence between the objective efficiency as measured in the experiment (i.e. Figure 5, top)

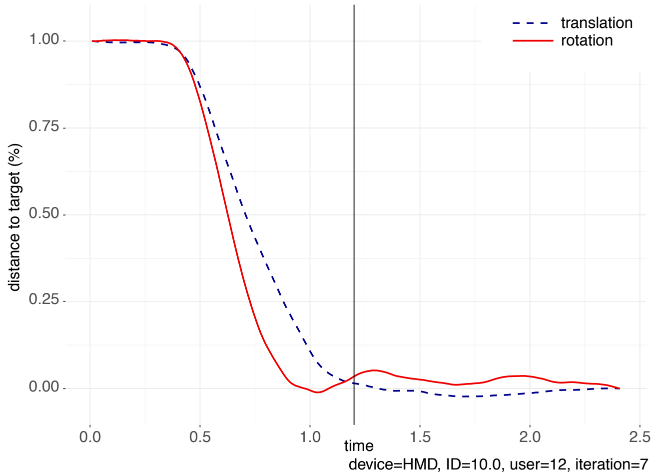

Figure 8. Error correction as a function of elapsed time. Trials are segmented (vertical line) in an initial ballistic phase followed by a precision phase.

and the subjective efficiency reported by the participants (Figure 7, efficiency). This tells that not only participants were more efficient with the HPCD, but also *felt* the benefit in performance. This is a good sign for the acceptability of this novel form of interaction technique. This may also explain why participant gave their best preference rating to this condition (Figure 7, preference).

Efficiency of the Large Sphere for Control

In two recent studies, the PHANTOM technique was found as the most efficient [19] and the most accurate [18] technique for docking. However, these studies did not include a large sphere held in two hands as an input device. In our experiment, the 3 conditions that used the large sphere yielded higher performances than the PHANTOM, indicating a better suitability of the sphere.

To analyze this result, we focus the discussion on the PROP and PHANTOM conditions that only differed by the input device. Overall, PROP yielded 18.8% better performances than PHANTOM. To get a more detailed picture of this result, we

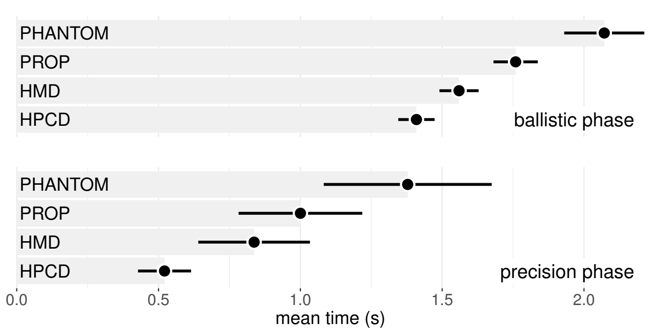

Figure 9. Time spent in the ballistic phase (top) vs. precision phase (bottom) of the dockings. Averaged across participants and ID with 95% confidence intervals.

analyzed the trajectory of the controlled object during the trials. Figure 8 illustrates a typical error correction trajectory for the translation and rotation parts of a docking. We segmented every trial into a ballistic phase followed by a precision phase: the threshold is chosen where the magnitude of the speed of change of the 2D error correction becomes less than 10% of its maximum. Figure 9 allows the comparison between conditions of the duration of the ballistic and precision phases.

It appears that PROP was more efficient than PHANTOM in both phases, although the benefit was more important in the precision phase. The PHANTOM benefitted from a smaller scale of the scene that allowed participants to quickly move the pen in the neighborhood of the target. However, the duration of the ballistic phase in PROP was still smaller, which indicates that covering a larger distance with the large sphere held in two hands was not an impediment. The large benefit of PROP in the precision phase can be attributed to the stability offered by holding of the sphere with 2 hands, to the large radius of the sphere that affords precise rotations, and to the lack of rotation constraints. In addition, we observed that 4 of the 20 participants required assistance when manipulating the pen in PHANTOM for the first time, while all users quickly managed to acquire the first target with the sphere. This indicates a more *intuitive* control of the sphere than of the pen.

Benefits of Co-Locating Motor Space and Display Space

Although using the same input devices in PROP, HMD and HPCD, participants were on average 12.4% and 29.5% more efficient with HMD and HPCD, respectively, than with PROP. HMD and HPCD both create the illusion to the user that the controlled object remains between their hands. This confirms a previous result from Ware and Rose who observed that "having the hand physically in the same location as the virtual object being manipulated" was an important factor for an efficient rotation of virtual objects with physical handles [21].

Superiority of HPCD over HMD

The HPCD and HMD condition used the same input device and superposed the scene in the motor and display spaces in a 1 to 1 mapping. In other words, participants had to perform the exact same sequence of gestures in the two conditions. Yet, participants' docking performances were quite different with HPCD yielding 17.0% better performances than HMD. To interpret this result, we analyze the way that the visual feedback was provided to the participants, as this was the only difference between the two conditions.

HPCD did have a number of drawbacks compared to HMD, but the higher achieved performances indicate that they were not important factors to execute the task efficiently.

- Participants could hide the projection with their head, but they quickly found a posture where this was not an issue.
- The limit of the projection area could have been another issue, but we chose a set of target position so that this was not the case during the experiment.
- The HPCD only offers a limited view of the scene through the sphere. However, as we were interested in the benefit of the proprioception, we used a limited number of targets

that were always in the same location, and we used the pointing arrow to help participants at the beginning of the session. Doing so, we removed the problem of *discovering the scene*.

One factor explaining the lower performance of HMD could be the lower pixel density in the focus area of the participants. However, this did not prevent HMD to yield higher performances than PROP; which had a similar pixel density than HPCD. As a consequence, we don't think that the pixel density was a major factor, although it will be interesting to test this effect when resolution is improved on future HMDs. The HMD did introduce some discomfort to some participants, as indicated on Figure 7 (pain). This is a well-known problem with opaque HMDs: they are very well tolerated by some users while others quickly get discomfort. In contrast, no participants complained with the HPCD.

Apart from the differences in the presentation of the graphical feedback, it is important to note that HPCD and HMD differ radically in the contextual information that is visually presented to their users. With HPCD, the virtual chair and targets appear only as a small part of a larger physical scene including the sphere, the user's hands and forearms, and the room and furniture surrounding her. With HMD, the visual context must be created and is thus much simpler. In our experiment, we simply added a textured ground plane. Participants expressed that they felt cut from the real world. This difference does not appear as an important factor on the initial ballistic motion; which benefits from the proprioceptive knowledge of the target poses and could be executed blindly. Indeed, Figure 9 reveals that most of the benefit of HPCD over HMD was on the precision phase of the dockings. Here, we assume that the better contextual information, in particular the vision of the sphere and the hands, may have provided a higher quality visual feedback than the simple digital chair in the control of the precise and accurate rotations. Indeed, a small mistake in the control of a rotation has a much more perceptible effect on objects that are far from the pivot, i.e. the surface of the sphere and the hands, than on the small chair centered in the sphere. The benefit of perceiving the physical surrounding is coherent with a previous study by Krichenbauer et al. who compared task performance in a VR and AR setup and found a 22% performance advantage of AR on a 9 dof task [9].

We estimated the coordination of participants' control by computing the m-metric on the object's recorded trajectories. The m-metric measures both the simultaneity of control and the efficiency of the coordination [11]. It has values in the range 0 (no coordination or total innefficiency) to 1 (total coordination with optimal efficiency). While coordination was similar in the ballistic phase (0.69 for both HMD and HPCD), the control was much more coordinated with HPCD in the precision phase (0.41 for HMD, 0.72 for HPCD). This confirmed that most of the benefit of HPCD over HMD was in the precision phase.

This large performance difference between HMD and HPCD is an intriguing result that will require further studies to fully understand its causes. One course of study would be to enrich the contextual information provided in the HMD and assess its effect on performances.

In any cases, this study indicates that an HPCD can provide a very efficient interaction technique for docking, especially when users can rely on a proprioceptive knowledge of the target poses. Further studies should test if the excellent performance remains when targets are not well known.

CONCLUSION

We tested 4 different isomorphic interaction techniques for 6-dof docking. The experiment was designed with the use of an Index of Difficulty (ID) for docking that we introduced. We observed a strong linear relationship between the task accomplishment time and ID; which supports the validity of the ID. This allowed the computation of an information throughput for each of the tested techniques. Our approach should facilitate the comparison of results across studies.

The experiment included the first measure of the performance of a handheld perspective coupled display (HPCD) for docking. The HPCD was found more efficient than all the other techniques. In particular, it was found 17% more efficient than an technique that used an HMD for display but exactly the same input as the HPCD. Most of the gain occurred in the final precision phase of the docking. This indicates that the rich visual cues offered by the physical world with an HPCD may be of particular importance for the precise and accurate 6-dof control. We intend to test this hypothesis in future work, for example by controlling the addition of several visual cues in the HMD condition, including a representation of the user's arms, hands and fingers. As a side effect, this could provide guidelines to improve docking techniques with HMDs.

ACKNOWLEDGEMENTS

This work was supported by the ISAR project ANR-14-CE24-0013.

REFERENCES

1. Bérard, F., Ip, J., Benovoy, M., El-Shimy, D., Blum, J. R., and Cooperstock, J. R. Did "minority report" get it wrong? superiority of the mouse over 3d input devices in a 3d placement task. In *IFIP Conference on Human-Computer Interaction, INTERACT*, vol. 5727 of *Lecture Notes in Computer Science*. Springer, 2009, ch. 45, 400–414.

2. Bérard, F., and Louis, T. The object inside: Assessing 3d examination with a spherical handheld perspective-corrected display. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2017), 4396–4404.

3. Bérard, F., and Rochet-Capellan, A. The transfer of learning as hci similarity: Towards an objective assessment of the sensory-motor basis of naturalness. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2015), 1315–1324.

4. Besançon, L., Issartel, P., Ammi, M., and Isenberg, T. Mouse, tactile, and tangible input for 3d manipulation.

In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2017), 4727–4740.

5. Bowman, D. A., Kruijff, E., LaViola, J. J., and Poupyrev, I. *3D User Interfaces: Theory and Practice*. Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 2004.
6. Froehlich, B., Hochstrate, J., Skuk, V., and Huckauf, A. The globefish and the globemouse: two new six degree of freedom input devices for graphics applications. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2006), 191–199.
7. Hancock, M., Carpendale, S., and Cockburn, A. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2007), 1147–1156.
8. Kin, K., Miller, T., Bollensdorff, B., DeRose, T., Hartmann, B., and Agrawala, M. Eden: a professional multitouch tool for constructing virtual organic environments. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2011), 1343–1352.
9. Krichenbauer, M., Yamamoto, G., Taketomi, T., Sandor, C., and Kato, H. Augmented reality vs virtual reality for 3d object manipulation. *IEEE Transactions on Visualization and Computer Graphics PP*, 99 (2017), 1–1.
10. Liu, J., Au, O. K.-C., Fu, H., and Tai, C.-L. Two-finger gestures for 6dof manipulation of 3d objects. *Comp. Graph. Forum 31*, 7pt1 (Sept. 2012), 2047–2055.
11. Masliah, M. R., and Milgram, P. Measuring the allocation of control in a 6 degree-of-freedom docking experiment. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (New York, NY, USA, 2000), 25–32.
12. Mine, M. R., Frederick P. Brooks, J., and Sequin, C. H. Moving objects in space: exploiting proprioception in virtual-environment interaction. In *ACM Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM Press/Addison-Wesley Publishing Co. (1997), 19–26.
13. Poupyrev, I., Weghorst, S., and Fels, S. Non-isomorphic 3d rotational techniques. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2000), 540–547.
14. Soukoreff, R. W., and MacKenzie, I. S. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International Journal of Human-Computer Studies 61*, 6 (2004), 751 – 789.
15. Spindler, M., Büschel, W., and Dachselt, R. Use your head: tangible windows for 3d information spaces in a tabletop environment. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS)*, ACM (2012), 245–254.
16. Stavness, I., Lam, B., and Fels, S. pcubee: A perspective-corrected handheld cubic display. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (2010), 1381–1390.
17. Stoelen, M. F., and Akin, D. L. Assessment of fitts' law for quantifying combined rotational and translational movements. *Human Factors 52*, 1 (2010), 63–77. PMID: 20653226.
18. Vuibert, V., Stuerzlinger, W., and Cooperstock, J. R. Evaluation of docking task performance using mid-air interaction techniques. In *ACM Symposium on Spatial User Interaction (SUI)*, ACM (2015), 44–52.
19. Wang, G., McGuffin, M., Bérard, F., and Cooperstock, J. Pop-up depth views for improving 3d target acquisition. In *Proceedings of Graphics Interface (GI)*, GI '11, Canadian Information Processing Society (2011), 41–48.
20. Ware, C. Using hand position for virtual object placement. *The Visual Computer 6*, 5 (1990), 245–253.
21. Ware, C., and Rose, J. Rotating virtual objects with real handles. *ACM Trans. Comput.-Hum. Interact. 6*, 2 (June 1999), 162–180.
22. Zhai, S., and Milgram, P. Quantifying coordination in multiple dof movement and its application to evaluating 6 dof input devices. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM Press/Addison-Wesley Publishing Co. (New York, NY, USA, 1998), 320–327.
23. Zhai, S., Milgram, P., and Buxton, W. The influence of muscle groups on performance of multiple degree-of-freedom input. In *ACM Conference on Human Factors in Computing Systems (CHI)*, ACM (1996), 308–315.